SUPPORTING INFORMATION APPENDIX

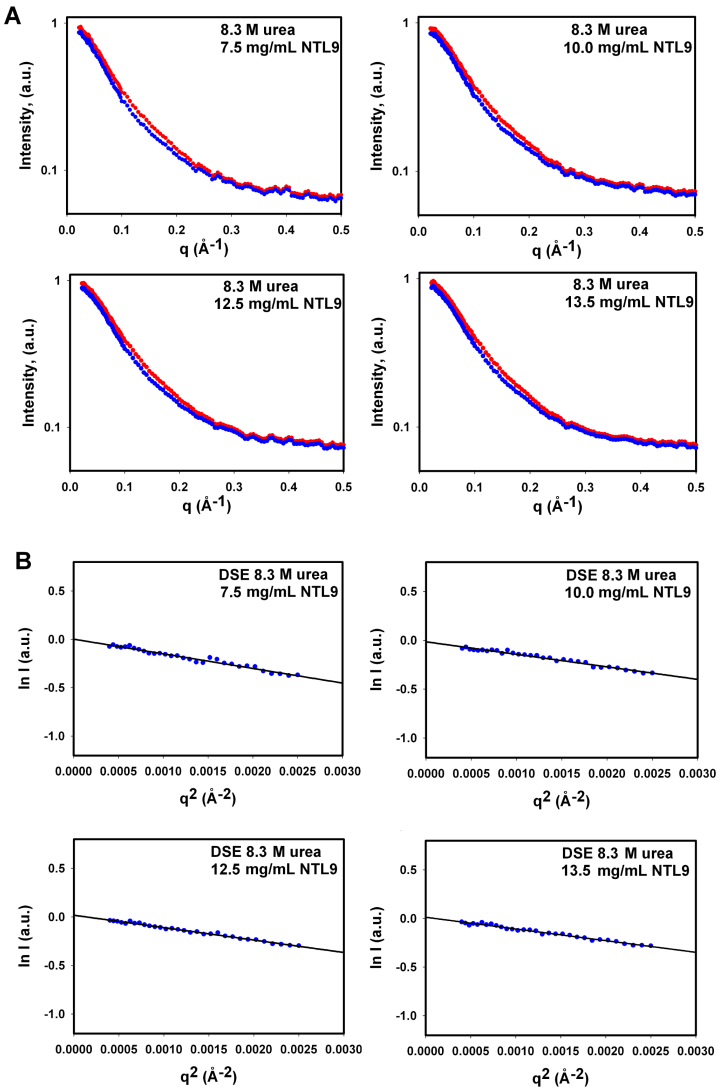

Fig. S1: SAXS determination of R_g for the urea unfolded state of NTL9 in 8.3M urea at 12°C, pH 5.5. 7% of the protein molecules are folded under these conditions and 93% unfolded. Experimental scattering profiles were collected for samples of NTL9 at 7.5, 10.0, 12.5, and 13.5 mg/mL in 8.3 M urea and native buffer. The population weighted contribution from the folded state was subtracted from the curves to obtain the scattering profile of the DSE in urea, and this data was used to estimate R_g. (A) Experimental scattering curves collected in urea (red) and curves after subtraction of the native state contribution (blue). (B) Guinier analysis of DSE scattering curves shown in panel A.

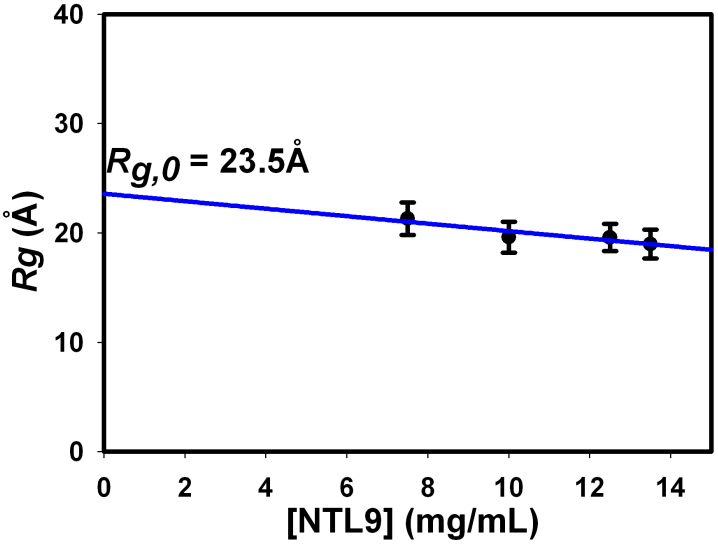

Fig. S2: Plot of the calculated DSE R_g as a function of protein concentration. The straight line is a linear fit to the data. The value extrapolated to zero concentration is 23.5 ± 3.5 Å. The uncertainty was estimated by bootstrap analysis.

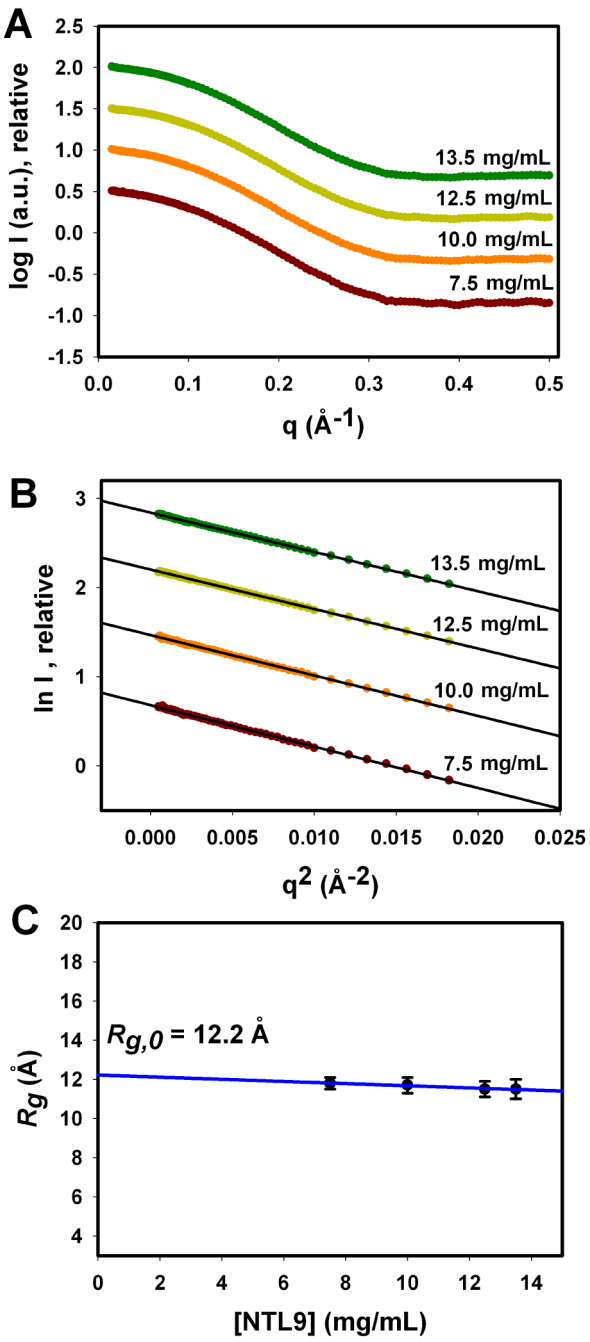

Fig. S3: SAXS profile for wild type NTL9 in native buffer at 12°C, pH 5.5. The protein is fully folded under these conditions. The curves are offset from each other for clarity. (A) The scattering curves. (B) Guinier analysis of data shown in panel A. (C) Plot of R_g as a function of protein concentration.

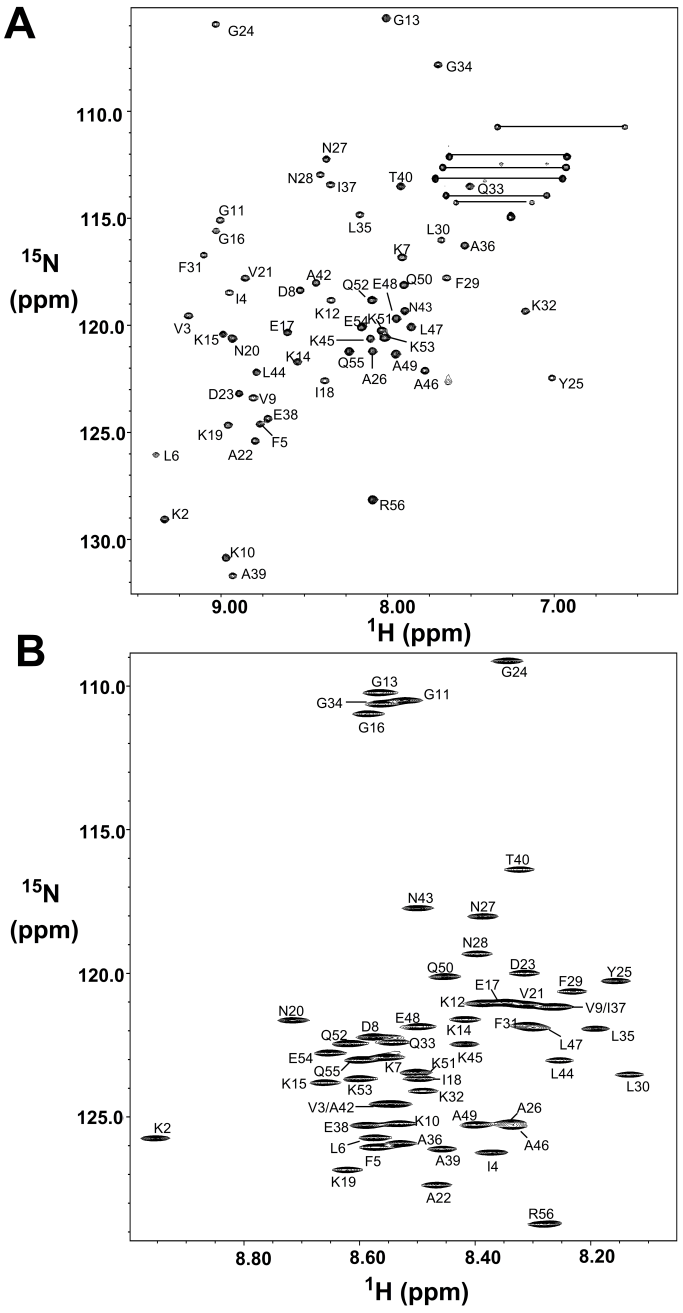

Fig. S4: Heteronuclear single quantum coherence (HSQC) spectrum of NTL9 under folding and denaturing conditions. (A) ^{1}H-^{15}N HSQC spectrum of NTL9 in the fully folded state in 90% H2O/10% D_2O, 20 mM sodium acetate, 100 mM sodium chloride, pH 5.5, 12 ˚C. (B) ^{1}H-^{15}N HSQC spectrum of NTL9 in 8.3 M urea, 20 mM sodium acetate, 100 mM sodium chloride, pH 5.5, 12 ˚C

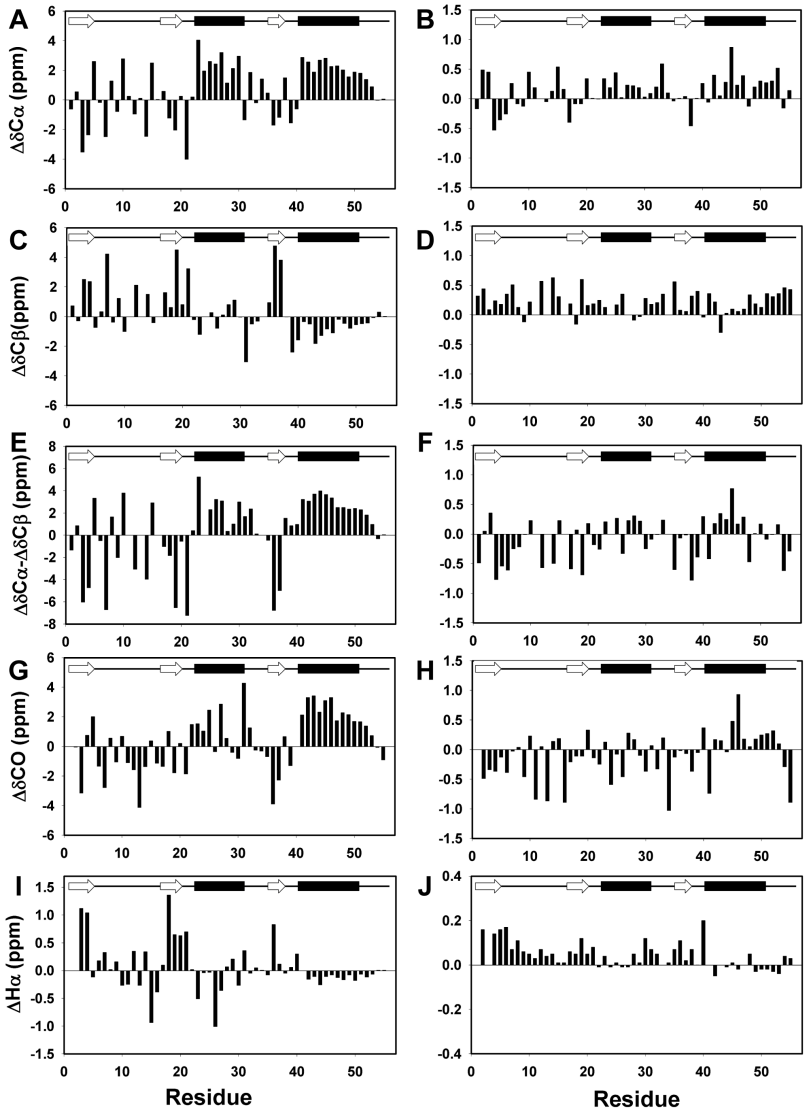

Fig. S5: Plots of secondary shifts for NTL9 in the fully folded state and DSE in 8.3M urea, pH 5.5, 12 °C. Data are plotted as observed minus random coil. (A) Folded $^{13}C_{\alpha}$, (B) DSE $^{13}C_{\alpha}$, (C) Folded $^{13}C_{\beta}$, (D) DSE $^{13}C_{\beta}$, (E) Folded $\Delta\delta^{13}C_{\alpha} - \Delta\delta^{13}C_{\beta}$, (F) DSE $\Delta\delta^{13}C_{\alpha} - \Delta\delta^{13}C_{\beta}$, (G) Folded ^{13}CO, (H) DSE ^{13}CO, (I) Folded $^{1}H_{\alpha}$ and (J) DSE $^{1}H_{\alpha}$. Random coil values of Wishart et al. were used (1). Note the different scales used for folded and DSE states. Schematic diagrams of the elements of secondary structure for the native state are shown at the top of each panel. β-strands are depicted as arrows and α-helices as black bars.

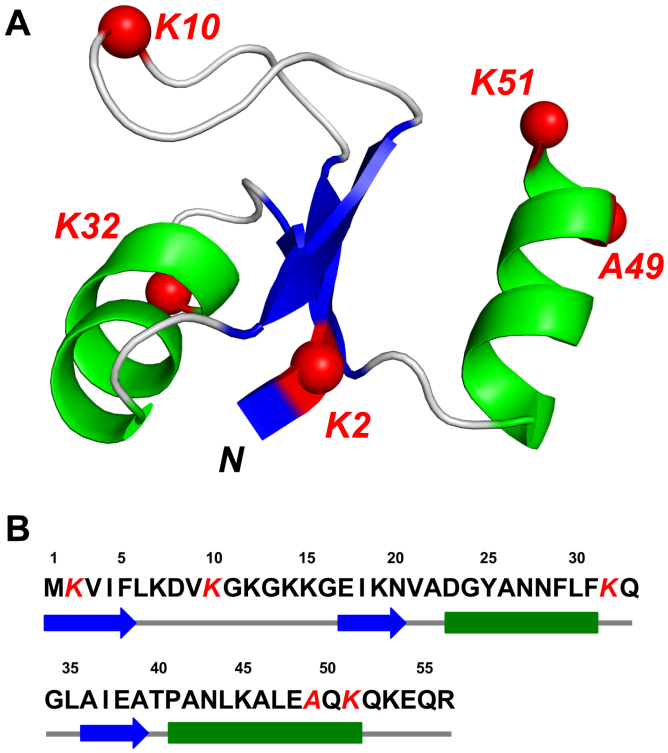

Fig. S6: Structure of NTL9 showing the spin label attachment sites. (A) Ribbon diagram of NTL9. The N-terminus is labeled N. The points of attachment of the spin labels are indicated as red spheres. (B) The primary sequence of NTL9 with a schematic diagram of the elements of secondary structure where arrows represent β-strands and bars represent α-helices. The sites of attachment of the spin labels are colored in red and are in italics.

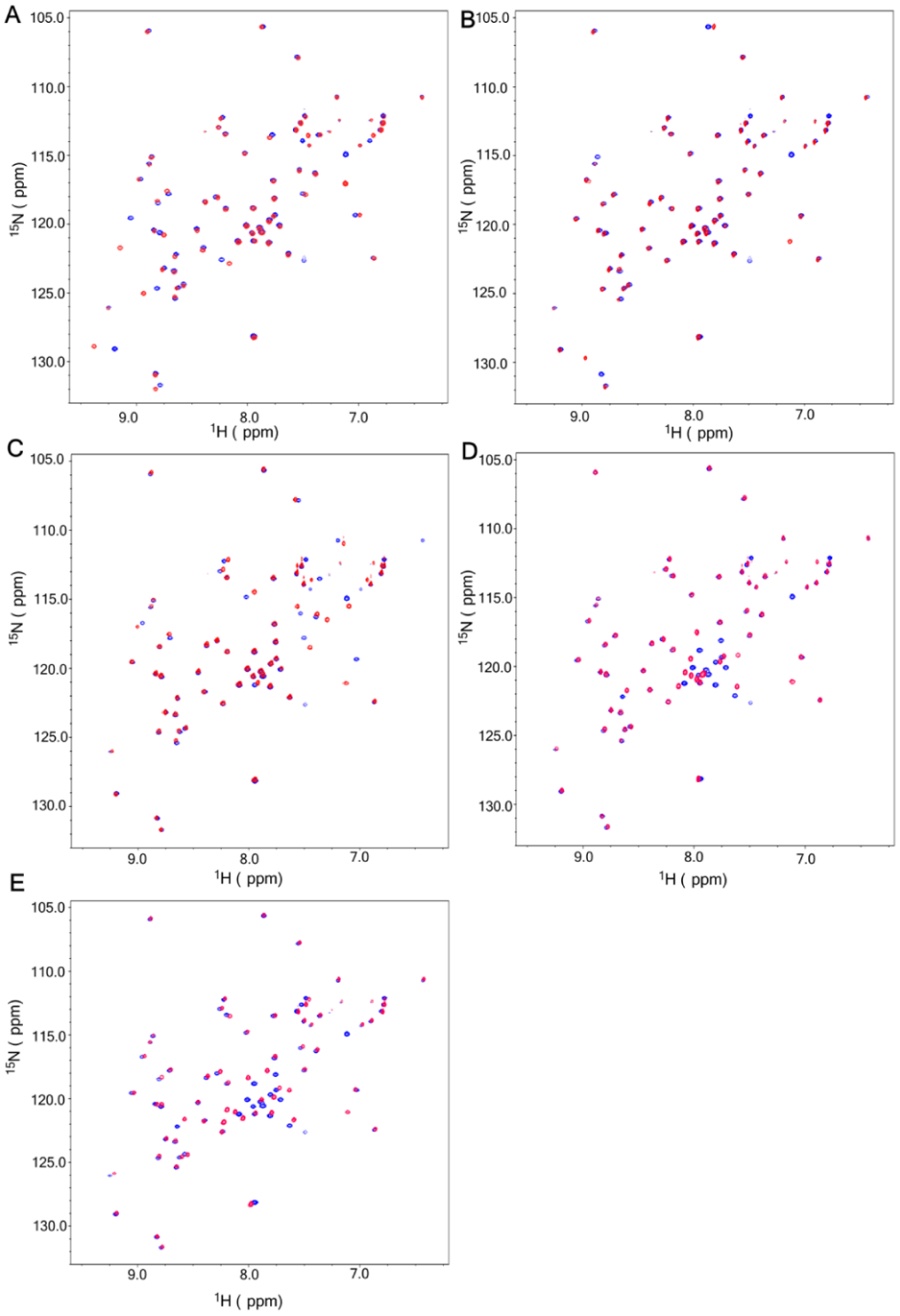

Fig. S7: Superimposed ^{1}H-^{15}N HSQC spectra of NTL9 wild type (blue) and its cysteine variants (red) for the folded state. (A) K2C, (B) K10C, (C) K32C, (D) A49C and (E) K51C. All NMR spectra were recorded at 12 ˚C, pH 5.5. Results of similar analysis for NTL9 in 8.3 M urea are shown in Fig. (S8).

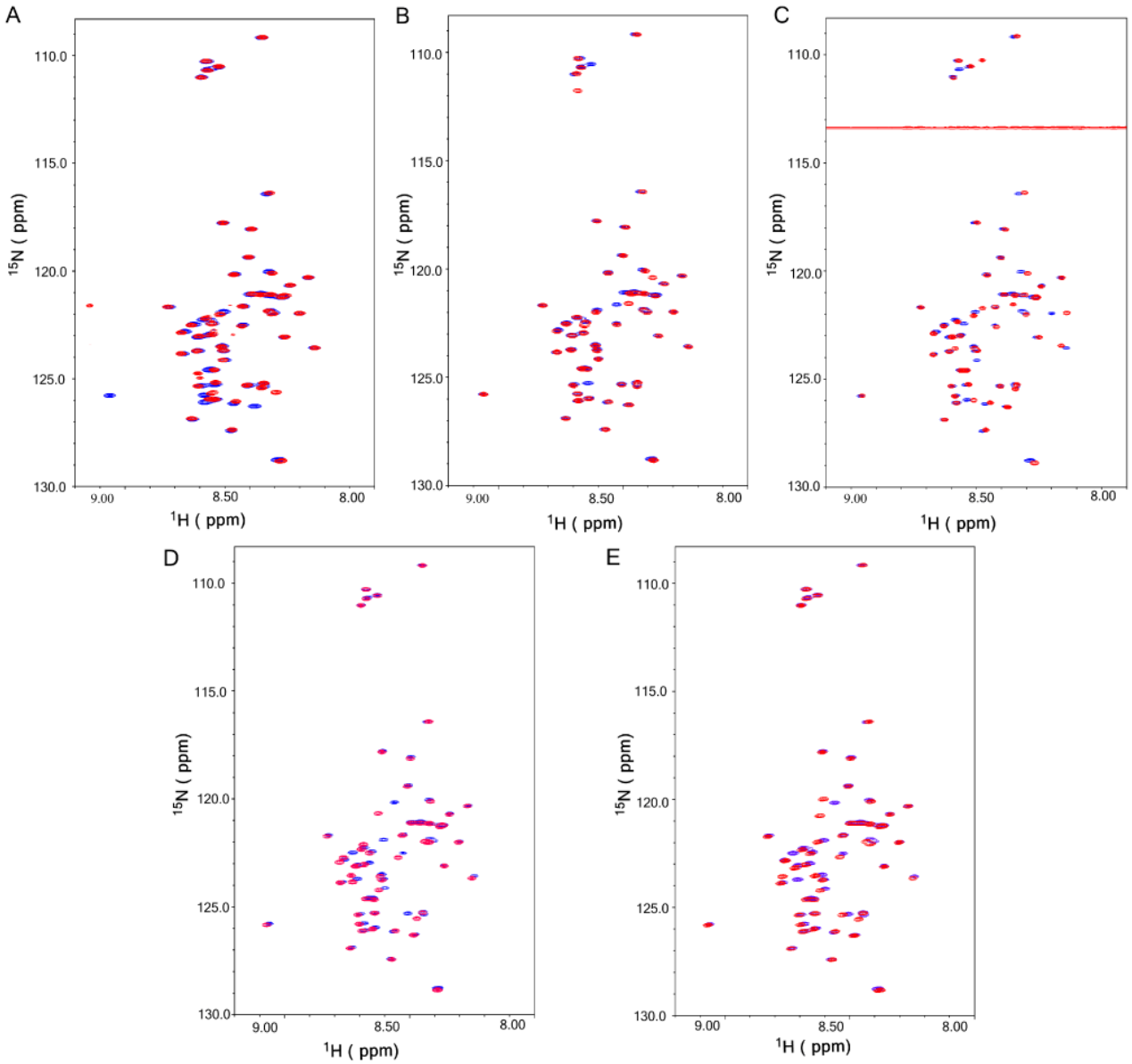

Fig. S8: Superimposed ^{1}H-^{15}N HSQC spectra of NTL9 wild type (blue) and its cysteine variants (red) for the urea denatured state. (A) K2C, (B) K10C, (C) K32C, (D) A49C and (E) K51C. All NMR spectra were recorded at 12 ˚C, pH 5.5, and 8.3 M urea.

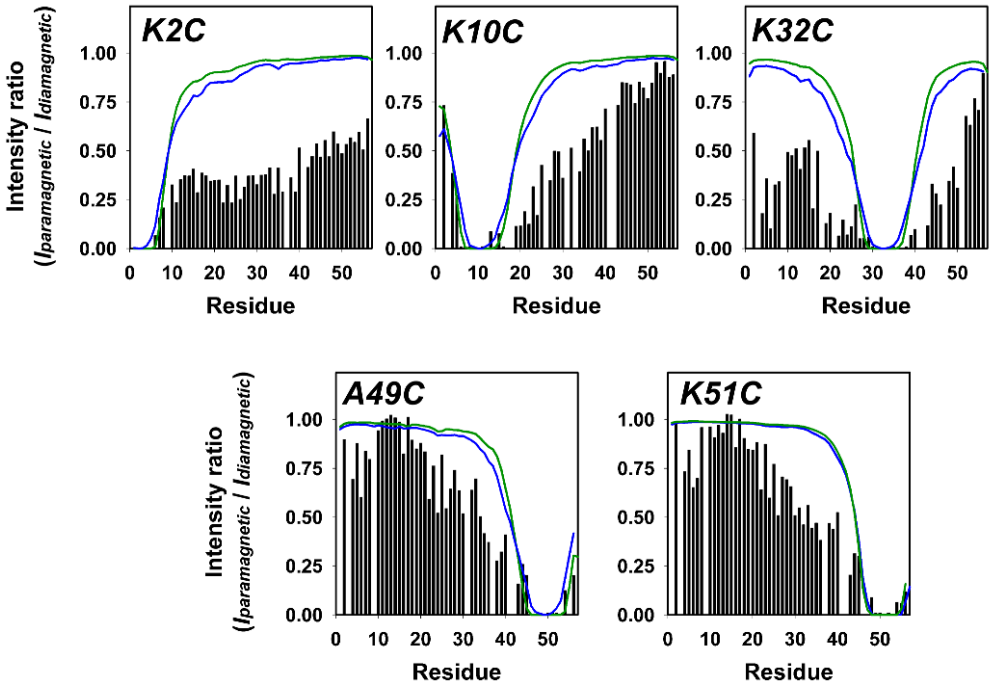

Fig. S9: Benchmarking models of the DSE. The expected PRE profile was calculated for each of the five spin labeled mutants using an excluded volume (EV) model ensemble that explicitly included the spin label (blue line) and that lacked the spin label (green line). For the calculation without the spin label the average distances were calculated from the C_β carbon of each Cys to the center of each amide NH bond in the protein backbone. The calculations including the spin label measured average distances from the nitrogen atom on the nitroxide to the center of each amide NH bond in the protein backbone.

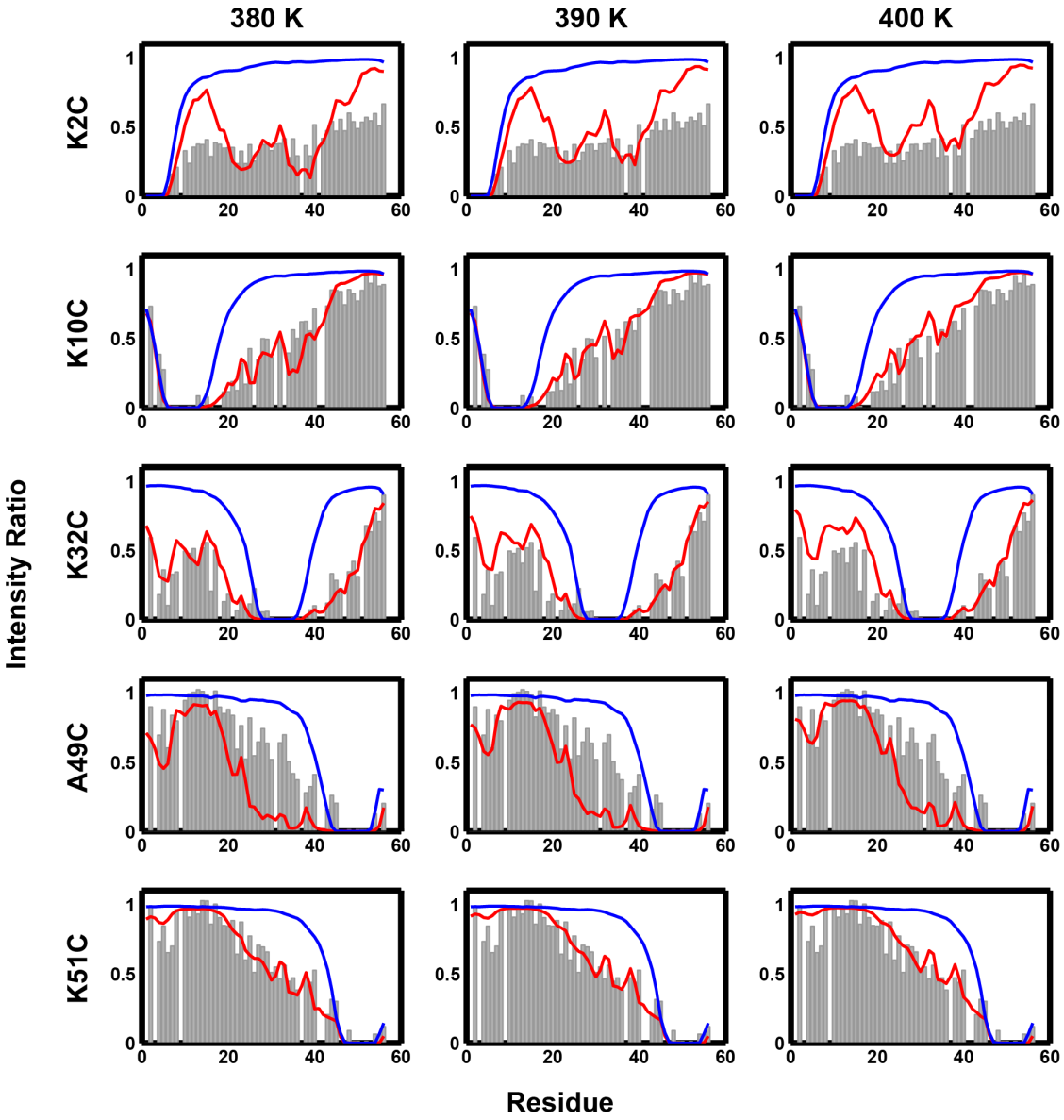

Fig. S10: Comparison between measured PRE data and calculated profiles. The gray bars denote experimental data for NTL9 in 8.3 M urea – shown here as peak intensity ratios between the diamagnetic and paramagnetic forms. The red curves are the profiles (intensity ratios) calculated using simulated ensembles at each of the T_D temperatures. Each column corresponds to a specific simulation temperature and each row to a specific NTL9 construct that is identified by the sequence position of the spin label in the experiment. In each panel, the blue curves denote the PREs calculated using EV ensembles for NTL9.

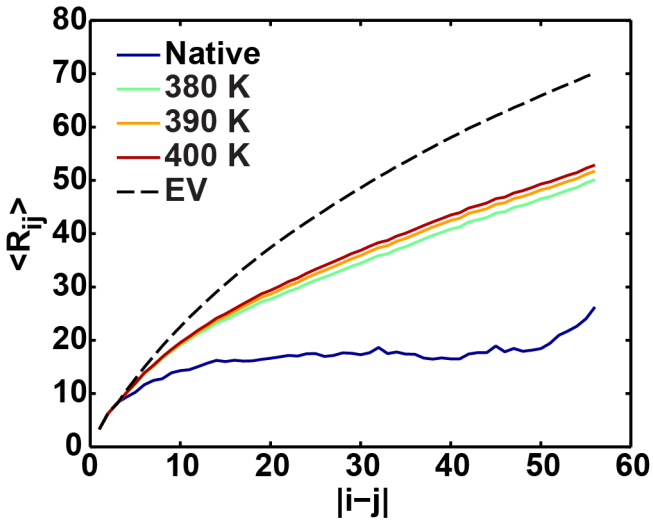

Fig. S11: Demonstration of the power-law scaling behavior of averaged inter-residue distances in the EV limit and T_D ensembles. The average spatial separations between residues i and j show power law dependence on sequence separation |i–j| whereas the folded ensemble (T = 240 Kelvin), which is a dominated by compact, globular conformations, shows different, plateauing behavior that is consistent with the density constraint imposed by chain compaction.

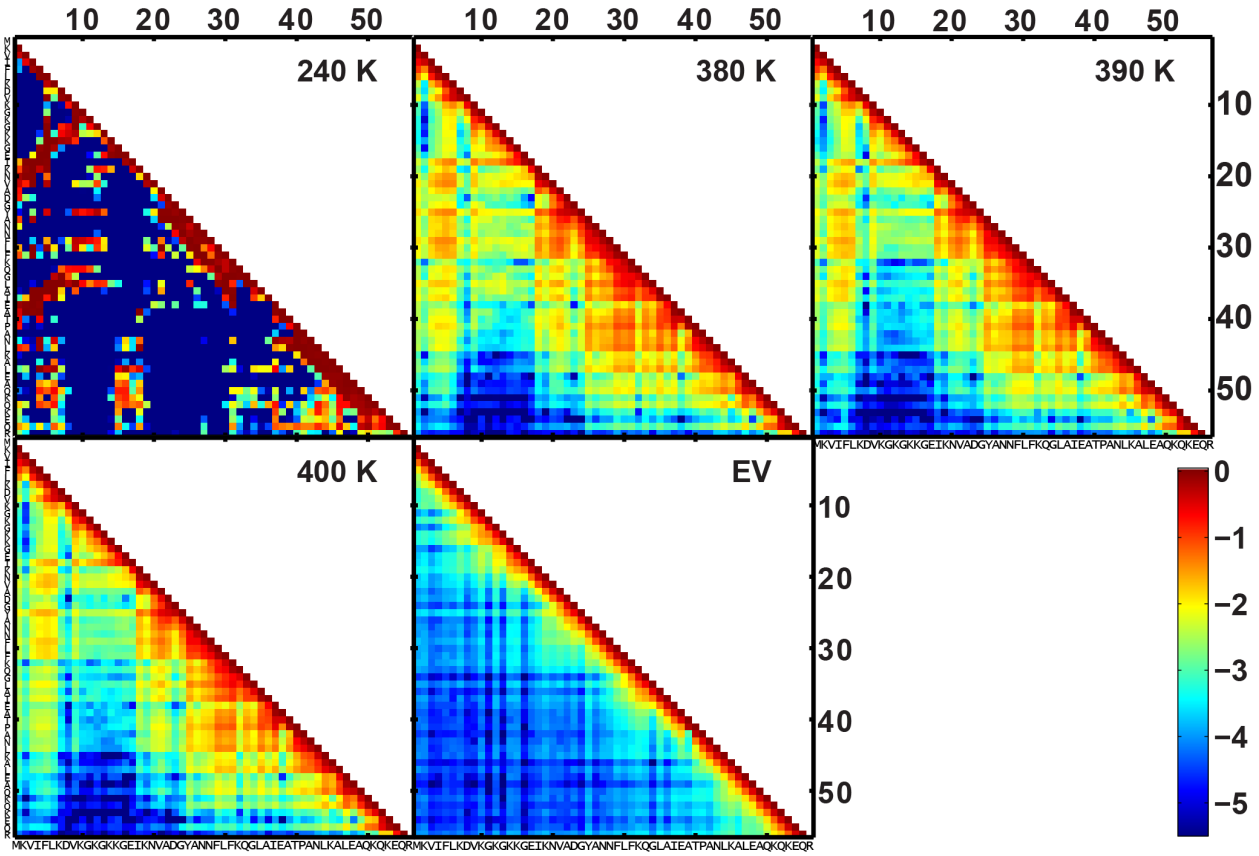

Fig. S12: Contact maps for the native state ensemble (240 Kelvin), the T_D temperatures, and the EV limit, respectively. In the contact maps, the coloring in each cell denotes the probability, p_{ij}, of realizing a contact between residues i and j. Only the lower triangular maps are shown because the contact maps are symmetric. To facilitate quantitative comparison between different ensembles, the plots show $\log_{10}(p_{ij})$ instead of p_{ij}. The same color bar is used for each contact map and this is shown in the bottom right corner.

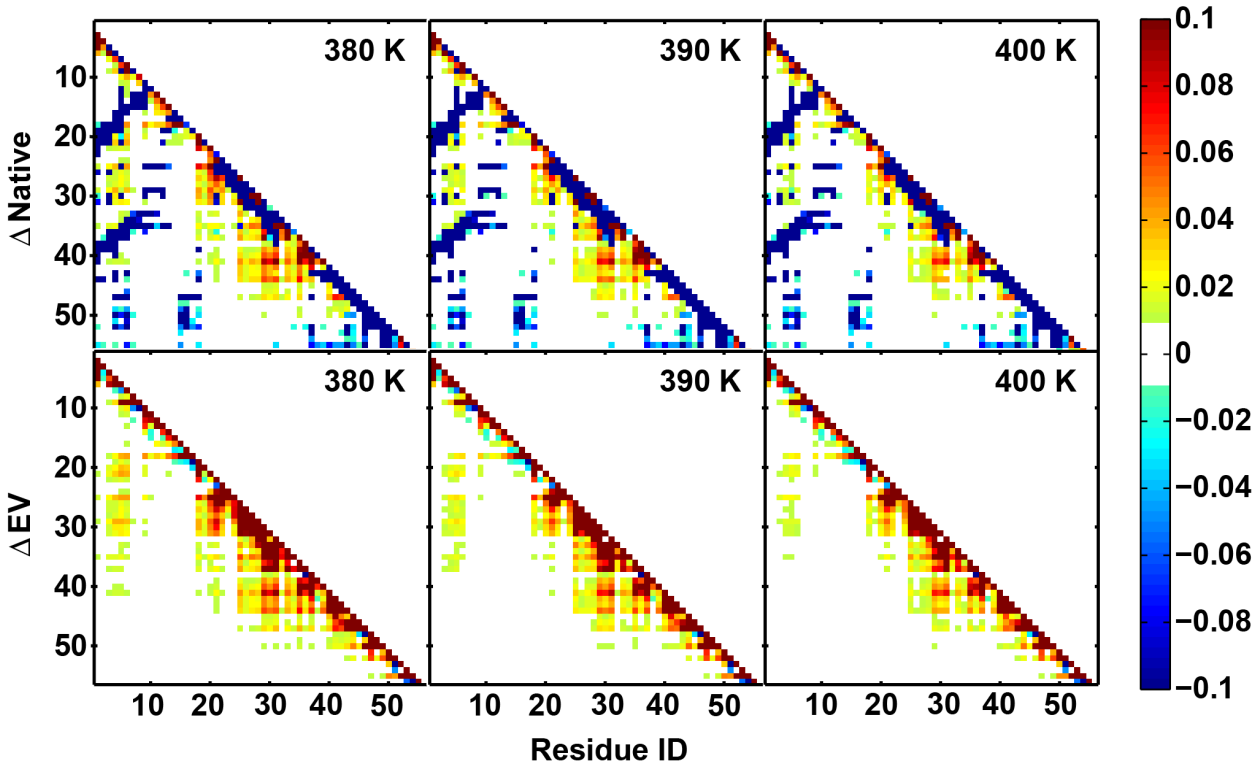

Fig. S13: Difference contact maps for each T_D temperature. The top row shows the difference contact maps with respect to the native state ensemble (240 Kelvin) and the bottom row shows similar maps with respect to the EV ensemble. For each T_D temperature the difference contact map is obtained by calculating $d_{ij} = p_{ij} - q_{ij}$, where p_{ij} is the average contact probability for a pair of residues i and j at a specific T_D temperature and q_{ij} is the probability for the corresponding contact in the reference ensemble (native or EV). In the difference maps, the cooler colors imply that the contact in question has a higher probability in the reference ensemble (native or EV) whereas the warmer colors imply that the contact has a higher probability in ensembles corresponding to the T_D temperatures when compared to the reference ensemble. A value of zero implies that the contacts either have similar probabilities in both the T_D and references ensembles or are missing in both.

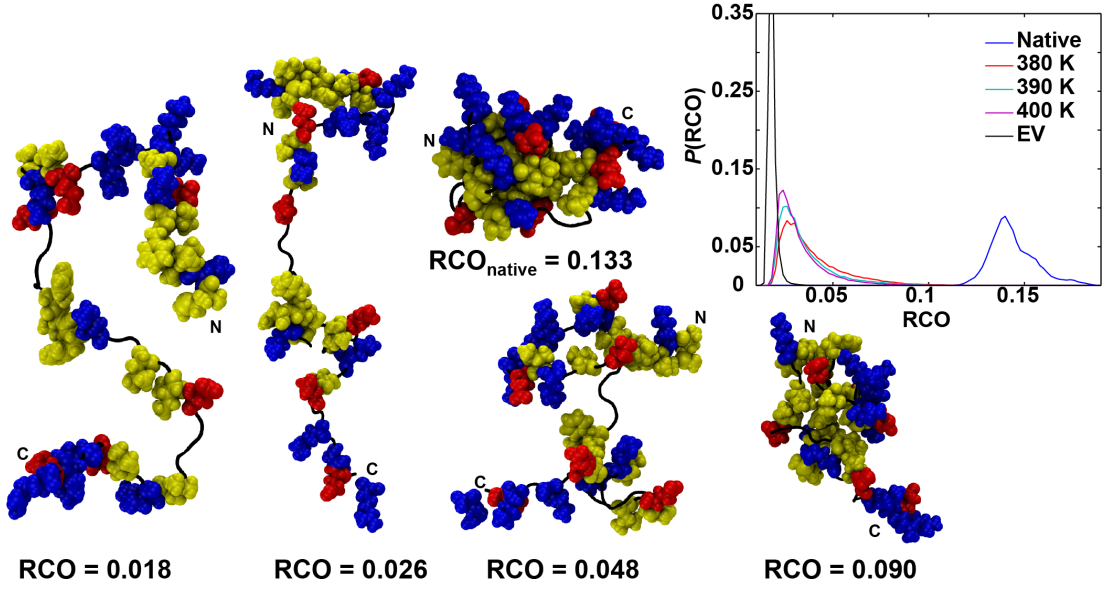

Fig. S14: Distributions (top right panel) of the relative contact order (RCO) for different ensembles and representative conformations drawn from the RCO distribution at 390 Kelvin (K). $\mathrm{RCO} = \frac{1}{N}\sum_{i=1}^{n_{\mathrm{contacts}}} s_i$; Here, N is the number of residues, n_{contacts} is the number of contacts in a specific conformation, and s_i is the sequence separation between the pair of residues that make up contact i. The bin widths are 0.002 RCO units, which are dimensionless. The overlap fractions between distributions for T_D temperatures and the EV and native state ensembles are as follows: (0.0935, 0.1136, 0.1325) and (0.0032, 0.0036, 8.33×10^{-4}, 1.83×10^{-4}) for 380 K, 390 K, and 400 K, respectively. In each snapshot, the polypeptide backbone is shown using a black contour. Positively charged sidechains are shown in blue, negatively charged sidechains in red, and hydrophobic groups in yellow. Each conformation drawn from the T=390 K ensemble is annotated by its corresponding RCO value *viz*., 0.018, 0.026, 0.048, and 0.090. The N- and C-termini are labeled in each snapshot. In addition, a snapshot is shown from the T=240 K native state ensemble with an RCO value of 0.133.

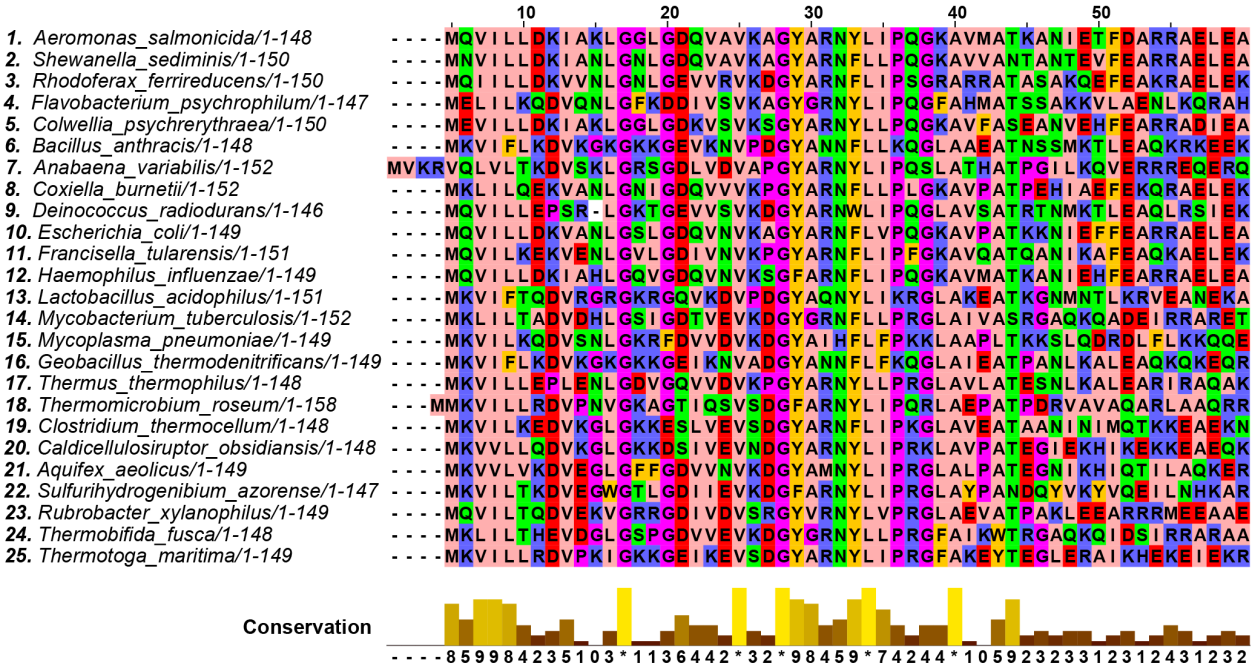

Fig. S15: Alignment of bacterial N-terminal NTL9 sequences. Sequences 1-5 are psychrophiles, 6-15 are mesophiles, and 16-25 are thermophiles. Multiple sequence alignment of 25 NTL9 variants was performed using version 2.6.1 of the Jalview software package. The color scheme used here is identical to the one used by the package ClustalX, which provides a graphical interface for the ClustalW algorithm. Each residue is assigned a color if the amino acid profile in the alignment meets a threshold criterion that is specific for the residue type. Positions in the alignment that do not match the threshold criterion are shown in white. The definitions of the ClustalW coloring scheme may be found at http://www.jalview.org/help/html/colourSchemes/clustal.html. The bottom two rows show the conservation scores and the consensus sequence, respectively. The former provides a numerical index that reflects the conservation of physico-chemical properties in the alignment. Identical residues score the highest on this scale followed by the next most conserved group, which contains substitutions to amino acids that belong to the same physico-chemical class.

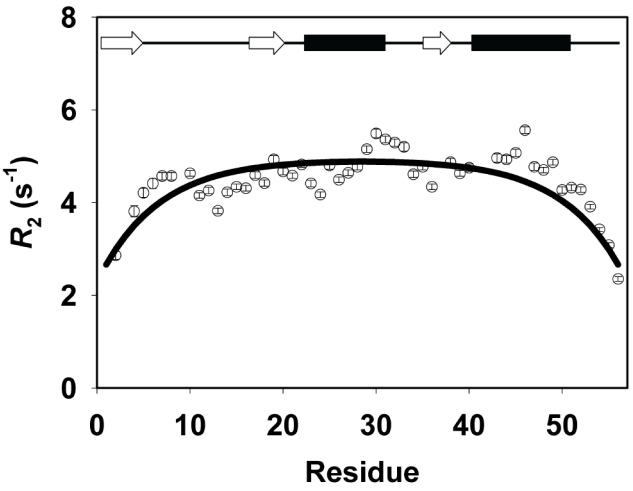

Fig. S16: ^{15}N R_2 rates for NTL9 in 8.3 M urea (○) at 12˚C, pH 5.5. The solid line is the best fit to the phenomenological model of Schwalbe and co-workers (2). Error bars are shown and are equal to or smaller than the size of the symbol. Uncertainties represent the apparent standard deviations determined using NMRViewJ (3) fitting routines. A schematic diagram of the elements of secondary structure of the native state is shown at the top of this figure.

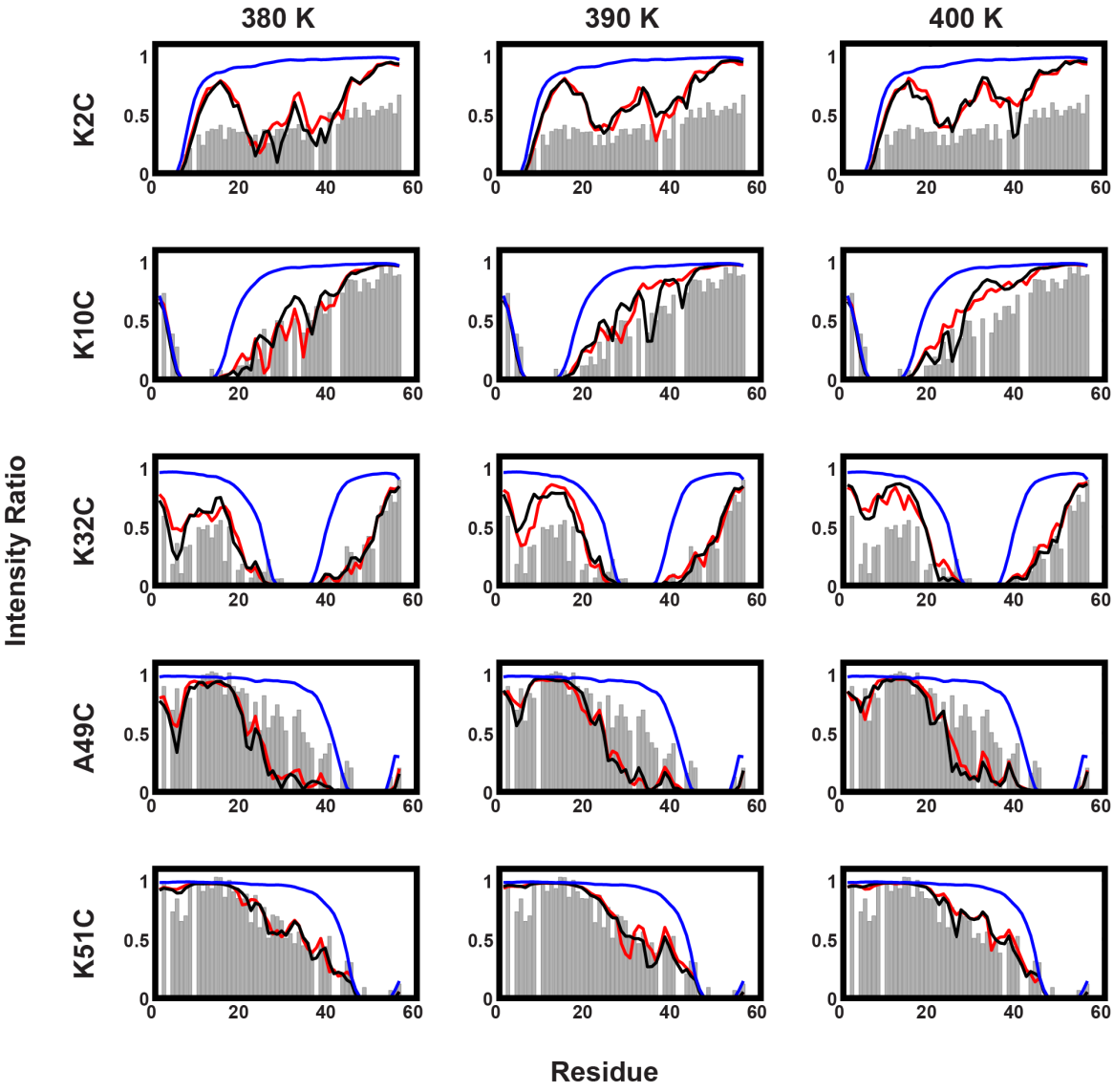

Fig. S17: Quantification of the dependence of calculated PRE profiles on salt (NaCl) concentration. The comparisons are shown in terms of the intensity ratios. The calculated profiles are as follows: EV limit (blue), 25 mM (red), and 120 mM (black).

Protein Expression and Purification. ^{15}N-labeled and ^{13}C,^{15}N-labeled NTL9 wild type and mutants were expressed in *Escherichia coli* BL21 cells in M9 minimal medium using standard methods. For ^{15}N-labeled proteins 0.8 g/L $^{15}NH_4Cl$ was used as the sole nitrogen source and for ^{13}C, ^{15}N-labeled protein 4 g/L [^{13}C] glucose was used as sole carbon source. The cells were grown at 37 ˚C until the OD at 600 nm reached 0.8, and then induced with 1 mM IPTG for 4 h. Cells were harvested and lysed by sonication. The proteins were purified with an ion-exchange column, and then by reverse-phase HPLC. The identity of each protein was confirmed by mass spectrometry.

Small Angle X-ray Scattering (SAXS) Experiments. Samples of NTL9 were prepared in native buffer (20 mM sodium acetate and 100 mM sodium chloride, pH 5.5), as well as in 8.3 M urea. Scattering experiments were performed at beamline X9 at Brookhaven National Laboratory, National Synchrotron Light Source I (Upton, New York, USA). Protein samples were injected into a 1 mm diameter capillary continuously during the measurement at a rate of 0.67 μL/s in order to avoid radiation damage. The exposure time for each measurement was 30 s. Scattering data was collected for four separate protein concentrations: 7.5, 10.0, 12.5, and 13.5 mg/mL at 12 ˚C. Each sample was measured three times and then averaged before data analysis. The program PRIMUS (4) was used for buffer subtraction, and the radius of gyration (R_g) was obtained using the Guinier approximation, Eq. (S1):

$$I(q) = I(0)\exp(-R_g^2 q^{2/3}) \quad \text{(S1)}$$

where $I(q)$ is the intensity at scattering vector q (5).

Estimation of R_g for the DSE of NTL9

We also estimated R_g for the DSE by directly fitting the experimental data. The observed R_g for a two-component system can be described as:

$$R^2_{g\ observed} = p_{native} R^2_{g\ native} + p_{DSE} R^2_{g\ DSE} \quad \text{(S2)}$$

where p is the population of the corresponding ensemble (6). Using the R_g for the native state, and the relative population of the native and DSE allows R_g of the DSE to be estimated. The analysis gives values of 20.7 Å for the 7.5 mg/mL sample and 19.3, 19.2, and 18.7 Å for the 10.0, 12.5, and 13.5 mg/mL samples. These values are in excellent agreement with the values calculated by subtracting the scattering profile of the native state from the observed profile. The values are summarized in Table S1.

Table S1. Estimated values of R_g for the DSE of NTL9.

Protein concentration (mg/mL)	Native state R_g (Å)	DSE R_g (Å) from analysis of Fig. S1	DSE R_g (Å) estimated from Eq. (S2)
7.5	11.8±0.3	21.3±1.5	20.7
10.0	11.7±0.3	19.6±1.4	19.3
12.5	11.5±0.4	19.6±1.2	19.2
13.5	11.5±0.5	19.0±1.3	18.7

NMR Sample Preparation. Samples for native state studies were prepared in 90% H_2O/10% D_2O with 20 mM sodium acetate and 100 mM sodium chloride, pH 5.5. For denatured state studies urea was added to a final concentration of 8.3 M determined by refractometry. All NMR experiments were recorded at 12 ˚C, and protein concentration of approximately 1 mM. 2,2-Dimethyl-2-silapentane-5-sulfonate sodium salt (DSS) was added as an internal reference (0.00 ppm) for all samples.

NMR Assignments. The ^{1}H offset frequency was centered at the water resonance and the ^{15}N offset frequency was set at 118.0 ppm. 2D ^{1}H-^{15}N HSQC, 3D HNCACB, CBCA(CO)NH, HNCO, and ^{1}H-^{15}N TOCSY-HSQC experiments were performed to generate assignments for the native state. 2D ^{1}H-^{15}N HSQC and 3D HNCACB, CBCA(CO)NH, HNCO, HNCA and ^{1}H-^{15}N TOCSY-HSQC were used to obtain the DSE assignments. Data was processed with the NMRPipe package (7), and chemical shift assignments were accomplished using NMRViewJ (3). ^{1}H chemical shifts were referenced to DSS directly, and ^{15}N, ^{13}C chemical shifts were indirectly referenced using standard methods. The random coil values of Wishart and coworkers were used to calculated secondary chemical shifts (1). Sequence-dependent corrections of random coil chemical shifts were performed (8).

Pulsed-field Gradient NMR Diffusion Experiments. A Bruker 600 MHz spectrometer equipped with a cryoprobe was used for the Pulsed-field Gradient NMR Diffusion experiments. Protein samples dissolved in deuterated urea were exchanged in 100% D_2O for 5 h at 25 ˚C and lyophilized. The exchange procedure was repeated three times to ensure that amide protons in the protein and the protons in urea were fully exchanged. The final concentration of urea was 8.3 M as measured by refractometry and the pH was adjusted to 5.1 (uncorrected pH meter reading). 1,4-dioxane was used as an internal standard. A pseudo-2D version of PFG-diffusion-ordered

spectroscopy (DOSY) was used for data collection and analysis. The diffusion delay, Δ, was set to 100 ms and the gradient pulse width, δ, was set to 8 ms. 32 spectra were collected for increasing gradient strengths from 2% to 95% of the maximum strength in a linear fashion. The resonance of Y25 in the unfolded state was used for the analysis. The reported hydrodynamic radius of dioxane, 2.12 Å, was used to calculate the hydrodynamic radius of the protein (9).

^{15}N R_2 Relaxation Experiments. ^{15}N R_2 relaxation experiments were conducted on ^{15}N-labeled wild type NTL9 in 8.3 M urea on a Varian 600 MHz spectrometer equipped with a cryoprobe using a Carr-Purcell-Meiboom-Gill (CPMG) sequence. The 180˚ pulse spacing in the CPMG sequence was 1 ms. Spectra were collected in an interleaved manner with relaxation delays set to 14 (×2), 28, 38, 50(×2), 62, 86(×2), 98 and 110 ms. The spectral width was 8012.8 Hz (^{1}H) × 1920.0 Hz (^{15}N) with 1024 (^{1}H) × 512 (^{15}N) complex points. A recycle delay of 3 s was used. R_2 rates were determined using NMRViewJ (3) to fit the peak intensities to two-parameter exponential decay. ^{15}N R_2 rates were analyzed using the phenomenological model of Schwalbe (2) by fitting the experimental R_2 rates to Eq. (S3).

$$R_2(i) = R_2(\text{int}) \sum_{j=1}^{N} \exp\left(-\frac{|i-j|}{\lambda}\right) \qquad \text{(S3)}$$

$R_2(i)$ is the experimental R_2 value for residue i, $R_2(int)$ is the intrinsic relaxation rate which depends on the temperature and the viscosity of the solution. N is the total number of residues in the protein, and λ is the apparent persistence length of the chain.

Preparation of Spin Labeled Samples. Samples were prepared by dissolving 3 mg of a NTL9 Cys mutant in 600 µL NMR buffer (20 mM sodium acetate, 100 mM sodium chloride). 12 µL of a 300 mM MTSL ((1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)methanesulfonate) stock

solution was added. The reaction was performed at room temperature for 10 h and a Sephadex G25 column was used to remove excessive MTSL. For native state experiments, the protein sample was split into two equal aliquots. For one of the aliquots, 300 μL NMR buffer was added (paramagnetic form); for the other aliquot, 270 μL of NMR buffer plus 30 μL of 100 mM TCEP (tris(2-carboxyethyl)phosphine) stock solution was added to reduce the protein (diamagnetic form) concentration. The pH was then adjusted to 5.5.

For the DSE studies, urea was added to a concentration of 10 M, (determined from the refractive index). The sample was split into two equal aliquots. For the diamagenetic form, 30 μL of 100 mM TCEP stock solution and NMR buffer was added; for the paramagnetic form, only NMR buffer was added. The final concentration of urea was 8.3 M for both samples, determined by measuring the refractive index.

Calculation of the Expected PRE Intensity Ratios and Γ_2 from the Analytical Gaussian Chain and EV Model. The Gaussian model assumes that the distance between each residue and the spin label site follows a Gaussian distribution for the root-mean-square distance between residues (10, 11):

$$< r^2 > = nl^2 \left(\frac{1+\alpha}{1-\alpha} - \frac{2\alpha(1-\alpha^n)}{n(1-\alpha)^2} \right) \tag{S4}$$

where r is the distance between a residue and the spin label site, n is the number of residues between residue i and the spin label site, l is the link length of the chain, taken to be 3.8 Å, and α is the cosine of the bond-angle supplements for the freely rotating chain model, which was set to 0.8 based on experimentally determined estimates of statistical segment lengths in poly-L-

alanine (10,11). The contribution of paramagnetic relaxation enhancement to the transverse relaxation rate, Γ_2 was calculated from Eq. (S5):

$$\Gamma_2 = \frac{K}{\langle r \rangle^6}(4\tau_c + \frac{3\tau_c}{1+\omega_H^2\tau_c^2}) \quad \text{(S5)}$$

Here, K is 1.23×10^{-32} cm^6s^{-2}, $<r>$ is the average distance between a given residue and the spin label site, ω_H is the proton Larmor frequency, and τ_C is the effective correlation time, which is 4 ns for NTL9 DSE in 8.3 M urea calculated from ^{15}N R_1 and R_2 relaxation rates. Values for $<r>$ were estimated from the EV simulations or calculated directly using the Gaussian chain model in Eq. (S4). The peak intensity ratios between the paramagnetic and diamagnetic forms were calculated using Eq. (S6):

$$\frac{I_P}{I_D} = \frac{R_{2D}\exp(-\Gamma_2 t)}{R_{2D}+\Gamma_2} \quad \text{(S6)}$$

Here, R_{2D} is the transverse relaxation rate of the backbone amide protons in the diamagnetic form of the DSE in 8.3 M urea. The average value was measured to be 14 s^{-1} using 1D NMR methods. The parameter t is the total duration of the INEPT delays, which is 12 ms for the HSQC pulse sequence.

Details of the Unfolding Metropolis Monte Carlo (MC) Simulations. The internal degrees of freedom included the backbone ϕ, ψ, ω and sidechain χ dihedral angles. Rigid-body moves simultaneously change rotational and translational degrees of freedom of the protein whereas translational moves were applied to alter the positions of mobile ions. The spatial cutoffs for Lennard–Jones and electrostatic interactions between net-neutral charge groups were set to 10 Å and 14 Å, respectively. No cutoffs were employed for computing the electrostatic interactions for

ions and sidechain moieties with a net charge. Sodium and chloride ions were modeled explicitly. The frequencies with which different moves were chosen are summarized in Table S2.

Table S2: Summary of MC move sets and frequencies used for all simulations[1]

Move type	Parameters
Rigid-body	9% (50%, 10Å, 20º)
Random cluster	1% (50%, 10Å, 20º)
Concerted rotation	6.3%
Omega (ω)	5.67% (90%, 5º)
Sidechain (χ_1, …, χ_n)	27% (4x, 60%, 30º)
Backbone (ϕ, ψ)	51.03% (70%, 10º)

[1]The first value listed in parentheses of row 2, column 2 and 3 of Table S2 is the fraction of moves assigned to finite perturbations whereas the remaining attempts fully randomize the corresponding degrees of freedom. The second and third values are the maximum displacements associated with translational and rotational moves for finite perturbations. Alterations to the ω angle involve random perturbations of randomly chosen angles. The two sets of values in parentheses of row 5, column 2 of Table S2 are the fraction of ω-moves that attempt a stepwise perturbation along with the maximum step-size. Sidechain moves perturb the χ-angles of a given sidechain in the peptide. In each attempt to alter sidechain degrees of the freedom, a random number of χ-angles are given random orientations. Sidechain moves are inexpensive and therefore several sidechains are sampled simultaneously during each move. The number of sidechains sampled is the first value in parentheses of row 6, column 2 of Table S2. The remaining two values in parentheses give the fraction of χ-moves with a finite perturbation and the maximum value of that perturbation. Backbone moves simultaneously perturb the ϕ- and the ψ-angle of a given residue. The values in parentheses are interpreted the same way as for ω-moves. Concerted rotations simultaneously perturb eight consecutive backbone dihedral angles using the algorithm developed by Dinner (12). This move set allows us to simultaneously probe multiple length scales simultaneously and efficiently.

Generation of the starting conformation used in all MC simulations. The bond lengths and bond angles were fixed at values prescribed by Engh and Huber (13). Steric overlaps in the starting conformation were removed using 10,000 MC steps of sampling at T = 260 Kelvin in the presence of harmonic torsional restraints that were applied over all torsional degrees of freedom with a force constant of 0.2 kcal / (mol-degree^2). The equilibrium values for each torsional restraint were based on the coordinates in 2HBB. The torsional restraints were then removed and the resulting structure was further refined using 1,000 steps of steepest decent

minimization and 1,000 steps of additional MC sampling at T = 260 Kelvin. The heavy-atom RMSD between the coordinates in 2HBB and the starting structure for all simulation temperatures was 0.93Å. The simulations in the EV limit were performed using protocols analogous to previously published work (14, 15).

For each temperature, we performed ten independent MC simulations. We used 22 simulation temperatures in all, leading to results from 220 independent simulations. The starting conformations were identical in each of the 220 simulations. For a specific temperature, the ten independent simulations used different random seeds. A total of 8×10^7 MC steps were used in each simulation and of these, the results from the first 2×10^7 steps were discarded as equilibration. Information regarding different observables was accumulated once every 5,000 MC steps. We used T-WHAM (16) to reweight and combine observables across all temperatures. The starting conformation for each of the 220 independent simulations was based on the coordinates deposited in the protein data bank, identifier 2HBB. In this model, the coordinates of five C-terminal residues were not resolved. These were constructed manually using the following procedure: The positions of all protein backbone atoms over the first 51 residues were constrained to their final position after the equilibration procedure described in the main text. All torsional degrees of freedom over the sidechains in the first 51 residues were also restrained as described in the main text. No constraints or restraints were applied over the last 5 residues and all backbone and sidechain degrees of freedom were sampled. The resulting system was subjected to 500,000 MC steps of sampling at T = 330 Kelvin followed by 1,000 steps of steepest descent minimization. The added segment adopts partially alpha-helical conformations and displays significant fraying of the C-terminal end. The heavy atom RMSD over the first 51

residues of 2HBB and the resulting structure was 1.14 Å. The final structure included all 56 residues and was used as the starting conformation for all MC simulations.

The Metropolis Monte Carlo sampling protocol. All results presented in this work were generated using the following thermal unfolding protocol: Ten independent MC simulations were used for each of the following temperatures T = {240, 260, 280, 290, 300, 310, 320, 330, 340, 345, 350, 355, 360, 365, 370, 375, 380, 390, 400, 430, 450, 500} Kelvin. Additional sampling enhancements such as temperature replica exchange were not used because 1) the quality of sampling was sufficient without its use and 2) since each simulation was truly independent from all others it allowed evaluation of the reproducibility of all observables across independent simulations.

Calculation of the Paramagnetic Relaxation Enhancement (PRE) intensity and Γ_2 profiles from MC simulations and comparison to experiment. The PRE intensity ratio ($I_{paramagnetic}$ / $I_{diamagnetic}$) and Γ_2 profiles were estimated from simulation for the following five sites in NTL9: K2, K10, K32, A49, and K51. Note that our simulations contain no mutations or spin labels on NTL9, which is in contrast to the experiments that include a nitroxide spin label and the EV simulations. The pairwise distance distribution between the β-carbon *i* located at each of the five PRE sites and each backbone amide nitrogen atom *k* was accumulated every 500 MC steps with a bin size of 0.2Å. T-WHAM (16) was used to re-weight and combine these distributions collected at all 22 temperatures to maximally inform the distribution at each single target temperature. The average inverse sixth power of each *ik* pair at each target temperature is calculated as follows:

$$\left[\left\langle \frac{1}{R_{ik}^6} \right\rangle_{T=T_{target}}\right] = \frac{\sum_{t=1}^{n_{temps}} \sum_{j=1}^{n_{bins}} \left(\frac{1}{r_j^{ik}}\right)^6 \omega_j^{ik}(T_{target}, t) P(r_j^{ik}) \Delta r^{ik}}{\sum_{t=1}^{n_{temps}} \sum_{j=1}^{n_{bins}} \omega_j^{ik}(T_{target}, t)} \quad \text{(S7)}$$

Here, $P(r_j^{ik})\Delta r^{ik}$ quantifies the probability that the distance r_j^{ik} assumes a value between $r_j^{ik} + \Delta r_j^{ik}$. In this case, $\Delta r_j^{ik} = 0.2$Å. The term $\omega_j^{ik}(T_{target}, t)$ is the T-WHAM derived weighting factor enabling the combination of distance distribution information across all simulated temperatures *t* to n_{temps} back to the ensemble at $T=T_{target}$. The contribution of paramagnetic relaxation enhancement to the transverse relaxation rate (Γ_2) was calculated in the same manner as shown in Eq. (S5):

$$\Gamma_2 = \frac{K}{\left\langle R_{ik}^6 \right\rangle}\left(4\tau_c + \frac{3\tau_c}{1+\omega_H^2\tau_c^2}\right) \quad \text{(S8)}$$

Eq. (S8) reduces to:

$$\Gamma_2 = 1.9745\times10^8 \left\langle \frac{1}{R_{ik}^6} \right\rangle \quad \text{(S9)}$$

The peak intensity ratios between the paramagnetic and diamagnetic forms were calculated using Eq. (S6).

These profiles were obtained for five constructs, which are distinguished by the positions of the spin labels namely: K2, K10, K32, A49, and K51. We compared the measured PRE data to profiles calculated from simulation results using the parameters Δ_1 and Δ_2, which are temperature-dependent root mean square deviations (RMSDs), defined in Eq. (S10).

$$\Delta_j = \sqrt{\frac{\left(\delta^T_{K2,j}\right)^2 + \left(\delta^T_{K10,j}\right)^2 + \left(\delta^T_{K32,j}\right)^2 + \left(\delta^T_{A49,j}\right)^2 + \left(\delta^T_{K51,j}\right)^2}{5}}$$

$$j = 1,2$$

$$\text{where } \delta^T_{i,1} = \sqrt{\frac{1}{N}\sum_{k=1,k\neq i}^{N}\left[\left\langle \frac{I_P}{I_D}\right\rangle_k^T - \left\langle \frac{I_P}{I_D}\right\rangle_k^{\text{expt.}}\right]^2}, \qquad \text{(S10)}$$

$$\delta^T_{i,2} = \sqrt{\frac{1}{N}\sum_{k=1,k\neq i}^{N}\left[\left\langle \Gamma_2\right\rangle_k^T - \left\langle \Gamma_2\right\rangle_k^{\text{expt.}}\right]^2}$$

$$\text{and } i \equiv \text{K2, K10, K32, A49, K51}$$

Here, N=56 denotes the number of residues in NTL9. For each construct, distinguished by the position of the spin label i, the experimental data provide estimates for Γ_2 and the intensity ratio $\left\langle \frac{I_P}{I_D}\right\rangle_j^{\text{expt.}} = \left\langle \frac{I_{\text{paramagnetic}}}{I_{\text{diamagnetic}}}\right\rangle$ at the location of each of the unlabeled residues, k. To calculate Δ_1 and Δ_2 we used the intensity ratio (j=1) and Γ_2 (j=2), respectively for each of the simulation temperatures, T.

Alignment of Bacterial N-terminal NTL9 Sequences. Full-length bacterial NTL9 sequences were located in UniProt (http://www.uniprot.org/). We sampled evenly for mesophiles, psychrophiles, and thermophiles. ClustalW (17) was used to align the full-length sequences and all settings were default: transition weighting = 0.5; alignment weight matrix = BLOSUM 62; gap opening penalty = 10; gap extension penalty = 0.1; end gap penalty = 0.5; gap distance = 1. Jalview (http://www.jalview.org/) (18) was used to display the N-terminal portion of this alignment.

References

1. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR* 5:67-81.

2. Klein-Seetharaman J, *et al.* (2002) Long-range interactions within a nonnative protein. S*cience* 295:1719-1722.

3. Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 278:313-352.

4. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, & Svergun DI (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 36:1277-1282.

5. Guinier A FB (1955) *Small-Angle X-ray Scattering* John Wiley, New York.

6. Choy WY, Mulder FA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Key LE (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. *J Mol Biol* 316: 101-112.

7. Delaglio F*, et al.* (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277-293.

8. Schwarzinger S*, et al.* (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* 123:2970-2978.

9. Wilkins DK, et al. (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry* 38:16424-16431.

10. Sung YH, Eliezer D (2007) Residual structure, backbone dynamics, and interactions within the synuclein family. *J Mol Biol* 372:689-707.

11. Lietzow MA, Jamin M, Dyson HJ, Wright PE (2002) Mapping long-range contacts in a highly unfolded protein. *J Mol Biol* 322:655-662.

12. Dinner AR (2000) Local deformations of polymers with nonplanar rigid main-chain internal coordinates. *J Comput. Chem* 21:1132-1144.

13. Engh RA & Huber R (1991) Accurate bond and angle parameters for X-ray structure refinement. *Acta Crystallogr.* A47:400.

14. Tran HT & Pappu RV (2006) Toward an accurate theoretical framework for describing ensembles for proteins under strongly denaturing conditions. *Biophysical Journal* 91(5):1868-1886.

15. Vitalis A & Pappu RV (2009) ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions. *J. Comput. Chem.* 30:673-699.

16. Chodera JD, Swope WC, Pitera JW, Seok C, Dill KA (2007) Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J Chem Theor Comput* 3:26-41.

17. Larkin MA*, et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.

18. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189-1191.